DOCUMENT RESUME

ED 159 194 TM 007 435

AUTHOR	Law, Alexander I.; Bronson, William H.
TITLE	Evaluation Trainer's Guide. The Evaluation Improvement Program.
INSTITUTION	Educational Testing Service, Princeton, N.J.
SPONS AGENCY	California State Dept. of Education, Sacramento.; Office of Education (DHEW), Washington, D.C.
PUB DATE	95
NOTE	88p.; For related documents, see ED 142 563 and 581
EDRS PRICE	MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS	Data Analysis; Data Collection; *Evaluation Methods; *Evaluators; Higher Education; Information Dissemination; Inservice Education; Measurement Goals; Measurement Techniques; *Planning; *Program Evaluation; Research Design; Research Utilization; Sampling; *Teaching Guides; Test Selection; *Training Techniques; Workshops
IDENTIFIERS	California Evaluation Improvement Project

ABSTRACT

This guide was designed for trainers presenting inservice workshops for program evaluators, teachers, principals, curriculum specialists, and others responsible for school programs. It is keyed to a companion volume, the Program Evaluator's Guide (ED 142 563), the primary source for workshop participants. Developed under the California Evaluation Improvement Project, both guides present detailed information concerning the purposes and processes of program evaluation, the role of the evaluator, and the development of an evaluation plan or design. Instruction is provided in selecting or developing assessment instruments, collecting and analyzing data, reporting evaluation results, and applying the findings. (CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED159194

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

MICROFICHE ONLY

Calif State
Dept of Educ.

EVALUATION TRAINER'S GUIDE

TM007 435

Evaluation Improvement Program

CALIFORNIA STATE DEPARTMENT OF EDUCATION
WILSON RILES, SUPERINTENDENT

EVALUATION TRAINER'S GUIDE

a manual developed as part of the California Evaluation Improvement Project under the leadership of the California State Department of Education, Wilson Riles, Superintendent of Public Instruction and Director of Education

Alexander I. Law, Chief
Office of Program Evaluation and Research

William H. Bronson, State Director
California Evaluation Improvement Project

The Evaluation Improvement Program

The activity which is the subject of this report was supported in whole or in part by the U.S. Office of Education, Department of Health, Education and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the U.S. Office of Education, and no official endorsement by the U.S. Office of Education should be inferred.

Copyright © 1977 by the California State Department of Education.
All rights reserved.

First Edition, June 1977

Published by Educational Testing Service, Princeton, New Jersey

Educational Testing Service, a nonprofit organization, is an Equal Opportunity Employer

FOREWORD

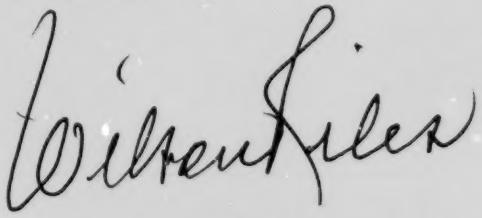
Evaluation of school programs is becoming more of a necessity for survival than a luxury enjoyed only by affluent districts. As financial resources diminish, decisions on how to allocate the available funds must be made. While basic educational research provides much valuable information, that information is usually not the kind on which day-to-day decisions about specific educational programs are based. Program evaluation, as perceived by the California Evaluation Improvement Project, is a means by which useful information is collected and analyzed by a local educational agency for its own use.

While most educators have had courses in testing and measurement and some contact with educational research, there has been little in their training to prepare them for conducting a systematic evaluation of a local school or classroom program. Of course, evaluation has been going on for many years, but it has most frequently been at the intuitive level, with little consistency and little impact on the total educational program.

California's response to this problem has been to develop a training program in basic evaluation concepts and skills, which is directed to the classroom teacher, the principal, the curriculum director, or program manager who wants to evaluate a local program to assist in local decision making.

One of the strengths of this training program is that it was developed and field tested throughout California by a group of educators whose backgrounds were primarily in the areas of program planning, curriculum, administration, and supervision. Evaluation specialists were used extensively as consultants as the workshop training program was developed, but the emphasis has been kept on how evaluation information could help in answering questions raised by the developers, whose orientation was basically that of program managers.

There is no magic formula to solve the problems involved in educating the youth of America, but I hope that this training program in basic evaluation concepts and skills will be useful to local schools and districts as they work toward improvement of the educational process.



WILSON RILES
Superintendent of Public
Instruction

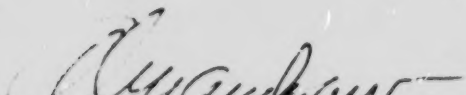
PREFACE

Program evaluation, through which a school or district evaluates its own program for its own purposes, is different from educational research. It is also different from a state testing program and from gathering information required by the state about achievement levels in specially funded programs. Program evaluation at the school or district level should be something the school or district does for itself for its own purposes rather than something an outsider does for it.

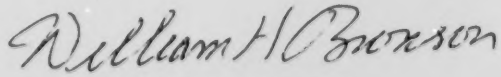
Program evaluation should be an integral part of the program-planning process. Provisions should be built into each program to collect information that will indicate progress towards the program's objectives, the degree of implementation of the plan, and other information required to make rational decisions about the program.

Program evaluation is of little value unless some use is made of its results. A part of the evaluation process includes identifying potential audiences for the evaluation report and finding out what kinds of information would be useful to them. Providing useful, timely information to people who can use it is one of the best ways of ensuring that the evaluation reports will be used.

These concepts are basic to the workshop materials that have been developed by the California Evaluation Improvement Project. The materials were designed to be as practical as possible for the educational practitioner, and it is our hope that the reader will find these concepts useful and will be able to apply them to future planning as well as to programs that are currently in operation.



Alexander I. Law
Chief, Office of Program
Evaluation and Research



William H. Bronson
State Director, California
Evaluation Improvement Project

INTRODUCTION TO THE EVALUATION IMPROVEMENT PROGRAM

Educational Testing Service (ETS) is pleased to have been selected by the California Department of Education as publisher, under an exclusive license, of the California Evaluation Improvement Project (EIP) materials. These constitute a course of instruction for individuals responsible for school programs and for those who help educational administrators ascertain program effectiveness. At the time of initial publication, spring 1977, the materials in EIP consist of the following:

- Program Evaluator's Guide. The Guide is a basic manual which provides in considerable detail background knowledge on the steps involved in planning and carrying out a program evaluation. It is designed as a study guide and learning tool for use in inservice training workshops for program evaluators.

- Workbook on Program Evaluation. The Workbook has two purposes. It can be put to use as a learning and instructional aid while one masters the procedures, techniques, and methods of program evaluation. Used this way, it helps the practitioner summarize and put into practice the subject matter presented in the Program Evaluator's Guide. It is best used, however, as a working notebook which the trained program evaluator can use for recording his or her plans as they are made and for making notes on program and program evaluation activities and events during the course of the program year. Used in this way, it helps the program evaluator keep complete records of the important information related to the program evaluation. It will probably be most useful when an interim or end-of-year program evaluation report has to be prepared, for much of the information needed at those critical times will already have been made a matter of record in the Workbook.

- Evaluation Trainer's Guide. This volume is a companion to the Program Evaluator's Guide. It supplies background and supporting materials for use by instructors conducting program evaluation workshops. Graphic art is provided for visual aids in support of a variety of subjects.

See page xiv for information about obtaining these materials.

Publication early in 1977 of the first of a continuing offering of EIP materials is consistent with ETS's long-term commitment to help advance the art of program evaluation in the elementary and secondary schools. The EIP materials are expected to go through a number of printings under the ETS imprimatur. Each successive printing will be a revised edition. Here we ask the help and cooperation of the readership.

As you, as a program evaluation practitioner, identify parts of any of these three works that could benefit from refinement and further development, or as you think of experiences that would serve to illustrate points made in any of the subject treatments, we hope that you will share your thoughts with us.

We would like to see the Evaluation Improvement Program subjected to its own program evaluation by those who use its materials. We hope the evaluation will be formative, not summative, in nature, for it is our intention to cycle evaluative comment on each edition into significant improvements in later ones. Present plans call for publication of the second edition, our first revision, late in 1977, and constructively critical comments from practitioners can be turned into refinements in print in very short order. Join with us to make the EIP materials, initially well developed by the California Evaluation Improvement Project, even better as time goes on. The California Department of Education and Educational Testing Service have joined in the common goal of making the EIP materials as practical and useful as they can be made to be.

Jack R. Childress
Vice President
Educational Testing Service

Wesley W. Walton
Program Director
Educational Testing Service

CONSULTANTS TO THE PROJECT

The following consultants worked closely with the developers of these materials and offered excellent suggestions which have been incorporated into the text:

Marvin C. Alkin, Ed.D., Director
Center for the Study of Evaluation
University of California at Los Angeles

Merlynn Bergen, Ph.D., Teaching and Research Assistant
School of Education, Educational Psychology
Stanford University, Stanford

Preston T. Bishop, Ed.D., Consultant
Division of Program Evaluation, Research and Pupil Services
Office of the Los Angeles County Superintendent of Schools

Antonio DePorcel, Ph.D., Senior Research Scientist
Behavioral Science and Technology Program
Ameican Institute for Research, Palo Alto

Annalee Elman, Ph.D., Research Assistant
School of Education, Educational Psychology
Stanford University, Stanford

J. Richard Harsh, M.A., Director
Los Angeles Office
Educational Testing Service

Roger A. Kaufman, Ph.D., Professor
United States International University

Larry E. Orcutt, Ph.D., Independent Consultant
L.E. Orcutt Associates, Incorporated

Dale M. Russell, Ed.D., Consultant
Division of Program Evaluation, Research and Pupil Services

R. Garry Shirts, Ph.D., Consultant
Simile II, Corporation

Les E. Shuck, Ed.D., Assistant Superintendent
Research and Development
Newport-Mesa Unified School District

Daniel L. Stufflebeam, Ph.D., Professor
Western Michigan University

Arlene B. Tenenbaum, Ph.D., Consultant
Educational Research Division
Xerox Corporation, Palo Alto

ADVISORY COMMITTEE, DIRECTORS AND STAFF

The following educators served as members of the California Evaluation Improvement Project State Advisory Committee and provided invaluable direction and support to the project:

Advisory Committee

- Robert W. Babcock, Ed.D., Director, Evaluation Improvement Center, Southern Section, Office of the Los Angeles County Superintendent of Schools
- Raymond M. Langley, M.A., Assistant Superintendent, San Luis Obispo County Office of Education
- Alexander I. Law, Ph.D., Chief, Office of Program Evaluation and Research, California State Department of Education
- Floyd I. Marchus, Ed.D., Superintendent, Contra Costa County Office of Education
- Donald A. MacLean, Ph.D., Assistant Superintendent, Orange County Office of Education
- Oliver "Bud" Neely, M.A., Assistant Superintendent, Shasta County Office of Education
- Nelson C. Price, Ed.D., Director, Evaluation Improvement Center, Northern Section, San Mateo County Office of Education
- William J. Zachmeier, Ph.D., Assistant Superintendent, Santa Cruz County Office of Education
- Donald C. Ziehl, Ed.D., Superintendent, La Canada Unified School District

State Department of Education

- Wlliam H. Bronson, M.A., State Project Director
- Carolyn M. Fowle, Ed.D., Consultant

Development Centers

- Robert W. Babcock, Ed.D., Director, Los Angeles County Office of Education
- John Plakos, M.S.
- Marie E. Plakos, Ed.D.

- Nelson C. Price, Ed.D., Director, San Mateo County Office of Education

- Patricia Evans, M.A.
- Carmen J. Finley, Ph.D.
- Arlen L. Kennedy, M.A.
- Alice W. Rotzel, Ph.D.

Satellite Centers

- Dean M Dennett, M.A., Director, Shasta County Office of Education

- Aniello L. Malvetti, M.A., Director, Sacramento County Office of Education

- Marti V. Halter, M.A.

- Phyllis L. McKinney, M.A., Director, Orange County Office of Education

- John Smith, M.A.
- Jeffrey A. Wells, M.A.

- Glen N. Pierson, Ph.D., Director, San Diego County Office of Education

- Rodney E. Phillips, Ph.D.

- Thomas J. Riley, Ph.D., Director, Fresno County Office of Education

- Jack M. Thompson, Ed.D., Director, Sonoma County Office of Education

- Vincent "Vic" Abata, M.A.
- Gregory A. Malone, M.A.

ACKNOWLEDGEMENTS

The initial materials were developed by the following EIP staff members:

- Robert W. Babcock, Ed.D., Director, Los Angeles Evaluation Improvement Project Development Center
- Patricia Evans, M.A., Consultant, San Mateo Evaluation Improvement Project Development Center
- Carmen J. Finley, Ph.D., Consultant, San Mateo Evaluation Improvement Project Development Center
- Arlen L. Kennedy, M.A., Consultant, San Mateo Evaluation Improvement Project Development Center
- John Plakos, M.S., Consultant, Los Angeles Evaluation Improvement Project Development Center
- Marie E. Plakos, Ed.D., Consultant, Los Angeles Evaluation Improvement Project Development Center
- Nelson C. Price, Ed.D., Director, San Mateo Evaluation Improvement Project Development Center
- Alice W. Rotzel, Ph.D., Consultant, San Mateo Evaluation Improvement Project Development Center

At the end of the first and second years of development, revisions were made by Carmen J. Finley, Carolyn M. Fowle, and William H. Bronson. Refinements were based upon extensive interviews with members of the EIP staff throughout the state who had been using the EIP materials in the conduct of workshops.

Development activities were coordinated by William H. Bronson, M.A., EIP Project Director, and Carolyn M. Fowle, Ed.D., Project Consultant.

Prepublication revisions were made by Wesley W. Walton, Ed.D., Director of the Evaluation Improvement Program at ETS. Nathaniel H. Hartshorne and Estelle Bartels served as ETS editors. Joan Westoff and Terry Birch provided covers and art supervision. Marissa G. Burch and Cathy E. Snyder served as text processing machine operators.

INFORMATION ABOUT EIP MATERIALS AND WORKSHOPS

Information about ordering Evaluation Improvement Program materials, about Evaluation Improvement Program workshops that use these materials, or about making arrangements for specially scheduled EIP workshops for local, regional, or state inservice training programs may be obtained by writing or telephoning the Evaluation Improvement Program at Educational Testing Service, Room P-069, Princeton, NJ 08540, (609) 921-9000 or at any of its regional offices listed below.

REGIONAL OFFICES OF EDUCATIONAL TESTING SERVICE

3445 Peachtree Road, NE
Suite 1040
Atlanta, Georgia 30326
(404) 262-7634

3724 Jefferson, Suite 100
Austin, Texas 78731
(512) 452-8817

1947 Center Street
Berkeley, California 94704
(415) 849-0950

2200 Merton Avenue
Room 216
Los Angeles, California 90041
(213) 254-5236

960 Grove Street
Evanston, Illinois 60201
(312) 869-7700

GPO Box 1271
San Juan, Puerto Rico 00936
(809) 763-3636, 3640, or 3670

One Dupont Circle
Suite 310
Washington, D.C. 20036
(202) 296-5930

2 Sun Life Executive Park
100 Worcester Road
Wellesley Hills, Massachusetts 02181
(617) 235-8861 or 8860

CONTENTS

GENERAL INSTRUCTIONS FOR THE EVALUATION TRAINER

Section A: PURPOSES AND REQUIREMENTS

Section B: PLANNING

Section C: DESIGN AND SAMPLING

Section D: INSTRUMENTS

Section E: DATA COLLECTION

Section F: DATA ANAYLSIS

Section G: REPORTING

Section H: APPLYING FINDINGS

EVALUATION TRAINER'S GUIDE

GENERAL INSTRUCTIONS FOR THE EVALUATION TRAINER

The Evaluation Improvement Program

GENERAL INSTRUCTIONS FOR THE EVALUATION TRAINER

Evaluation Improvement Program materials in program evaluation have worked well over the past three years in workshops of up to 30 persons and an instructional staff of two evaluation specialists.

An equally satisfactory arrangement has been to have an educator-generalist to assume responsibility for instruction in evaluation purposes, evaluation planning, data collection, and reporting and applying evaluation results as well as an evaluation specialist to assume responsibility for instruction in evaluation design, instruments, and data analysis.

Where there are more than 30 participants it has been necessary to divide the participants into two groups and to provide an instructional staff of four. With two groups, it is possible to have dual tracks, one for those with substantial backgrounds in quantitative methods, the other for those with less knowledge of statistics, sampling, and design.

The importance of carrying out the learning exercise cannot be emphasized too strongly. The exercises provide welcome opportunities for workshop participants to interact with the subject matter and with each other. So long as the exercises move along at a good rate of speed, they add a great deal to the "take home" value of the workshop. Since most exercises are group activities, workshop space should be arranged whenever possible with three to five persons at each table. Six groups of five persons each has proven to be an effective way to promote good small group interaction and sharing of findings.

The nature of the subject matter provides many fine opportunities for lively discussions on sometimes controversial topics, discussions that can be sharpened by the instructor through carefully framed questions or sequences of questions. Successful discussion periods have proven to be a matter of balancing time spent on such motivational activity with time devoted to familiarization with and uses for the EIP subject matter. Other discussions will result from questions by participants. It is important to give painstaking attention to such questions, especially those that are of general interest. It should also be emphasized, however, that discussions of over a few minutes on points of clarification often cause a workshop to lag. Questions of a technical nature usually are best left for after the scheduled workshop session.

Although the Evaluation Trainer's Guide is keyed to the Program Evaluator's Guide, the Workbook on Program Evaluation has proven to be as useful in the teaching and learning process. Inspection of the latter volume will show that its organization is compatible with those of the other two volumes, and it is an easy matter to find in it material to reinforce any point that might arise.

Time spent in starting some of the Workbook activities during a workshop, for later completion at home, has been found to have been time well spent. A quick page-flipping review of the Workbook makes an excellent initial orientation or concluding summary.

The primary purpose of the Evaluation Trainer's Guide is to lend support to the trainer in his or her presentations of workshop material, with the Program Evaluator's Guide being the primary resource for participants. The Trainer's Guide is designed to be used at the lectern or head table. There are cross-references throughout to the appropriate pages of the Evaluator's Guide and to the Learning Exercises, which should obviate any searching for subjects. Material surrounded by boxes is suggested text for instructors' presentations. "Topics Covered " and "Things to Stress" should be reviewed before a presentation. "Suggestions for Teaching" can be followed in the course of a presentation.

For the most part, the sequence of instruction is in ascending order as one moves through the Evaluator's Guide. The text has been written to be used with a set of 103 transparencies. Those used with each section are indexed on the back of each section's title page. The transparencies are of illustrations, charts, graphs, and other art work reproduced in the Evaluator's Guide, and their page references are part of that index. They can be made into transparencies locally. It is expected nonetheless that the prepared transparencies will be published as part of the Evaluation Improvement Program and will be available for distribution in summer 1977. Facilities, materials, and equipment needed to conduct an EIP workshop include the following:

- A conference room large enough for 30 persons clustered in groups of not more than 5
- Table space for participants so they can spread out their papers and books
- One copy for each participant of the
  Program Evaluator's Guide
  Workbook on Program Evaluation
- One copy for each instructor of the above publications and the Evaluation Trainer's Guide
- Chalkboard or chart paper on easel
- Chalk or wide-tip felt pens
- Overhead projector and electrical connection
- Screen
- Blank acetates and transparency marking pens for participants to use in preparing their presentations
- Homemade transparencies of art work in Program Evaluator's Guide or prepared transparencies from the publisher

Please give your careful attention to the final paragraph on page viii. Send us your comments for improvement of the EIP materials. Constructive criticism from practitioners will be reflected as refinements in the publications in their next edition.

EVALUATION TRAINER'S GUIDE

Section A

PURPOSES AND REQUIREMENTS

The Evaluation Improvement Program

TRANSPARENCIES

Transparencies	Subjects	Trainer's Guide	Evaluator's Guide
TR. A-0	Workshop	A-1	
TR. A-1	Reading Scores	A-2	A-2
TR. A-2	Evaluation	A-3	A-5
TR. A-3	Plan-Conduct-Use	A-4	A-6
TR. A-4	The Evaluation Process	A-4	A-6
TR. A-5	Types of Evaluation Data	A-4	A-7
TR. A-6	Interdependence of Program Parts	A-5	A-9
TR. A-7	Planning	A-6	A-13
TR. A-8	Goals and Objectives	A-7	A-15
TR. A-9	Questions to Raise in Setting Objectives	A-7	A-15
TR. A-10	Program Objective	A-7	A-16
TR. A-11 TR. A-12 TR. A-13 TR. A-14 TR. A-15 TR. A-16 TR. A-17 TR. A-18 TR. A-19	Learning Exercise 3	A-7 A-8	A-29 Through A-38
TR. A-20	Workshop Review	A-8	A-16
TR. A-21	Key Questions to Consider	A-8	A-16, 17
TR. A-22	Plan Ahead	A-9	A-22

LEARNING EXERCISES

1.	Types of Evaluation Data	A-25
2.	Identification of Measurable Objectives	A-27
3.	Selecting Appropriate Objectives	A-29
4.	Matching Needs Statements to Program Objectives and Program Activities	A-39

Introduction

The purpose of this section is to present the varied purposes and requirements of evaluation as perceived by various audiences. It also gives a brief overview of the evaluation process with a delineation of the initial steps to be taken by the evaluator in the evaluation process.

Introduction to Workshop (at page A-1 of the Evaluator's Guide)

1. TR. A-0 Project transparency on screen while giving an introduction to the workshop. The trainer may develop the introduction as he wishes or use the comments in the boxes.

> The demand for evaluation of educational programs has greatly increased in recent years. The state legislatures have insisted upon evidence of achievement when special funding has been given to programs; the federal government makes similar demands for accountability when federal funds are utilized by local school districts, and local communities are becoming more and more vociferous in their desire to be apprised of demonstrated accomplishments of the local school programs.
>
> As more and more districts adopt austere budgets, each program must be carefully scrutinized as to its effectiveness. Some programs may be eliminated, others expanded, still others modified to take account of their proven accomplishments and observed deficiencies.
>
> The California Evaluation Improvement Project has been developed to assist program managers in making effective evaluation plans so that data may be properly collected, analyzed, and reported. The key goal of program evaluation is to provide good answers to the questions that legislators, the community, and educational decision makers are asking about programs carried out in the schools.

Purposes of Evaluation

2. TR. A-1 This transparency may be used to initiate a discussion of the various purposes and reasons for program evaluation. Possible questions for the trainer to ask after showing the transparency are as follows:

   - What do you think the public reaction might have been to this newspaper report?
   - Do you think the public had any other information regarding pupil achievement?
   - If there had been frequent and comprehensive reporting to the public, would the report have caused such an uproar?

   Emphasize that one reason for evaluation is to keep the public informed.

   Elicit other reasons or purposes for evaluation from the audience. Have these responses recorded on a chalkboard or easel paper in front of the room by one of the participants. Group the responses into four or more major categories of purposes for program evaluation, such as (1) communicating with the public; (2) providing information to decision makers; (3) improving existing programs; and (4) providing satisfaction to the participant. In some situations, the trainer may have to supplement the responses from the audience.

3. Move into a discussion that emphasizes the groups served by those purposes. Ask the participants to suggest groups for each. Examples are as follows:

   Communicating with the Public

   Parents
   Business community
   Professional community
   Industrial community
   Retired community
   Special interest groups

Providing Information for Decision Makers

Classroom teachers
Principal
District administrators
Superintendent
Board of education

Improving Existing Programs

Classroom teachers
Principal
District administrators
Superintendent
Board of education
Citizens advisory committee

Providing Satisfaction to Participants

Students
Classroom teachers
Instructional aides

Stress that each group has its own questions and therefore may require different information from the program evaluation.

4. Ask the participants to take a few minutes to read Purposes of Program Evaluation on pages A-1 through A-5 of the Guide. This will afford review and reinforcement of the preceding discussion.

Overview of the Evaluation Process (at page A-5 of the Guide)

1. TR. A-2 Project transparency while defining program evaluation.

> Program Evaluation, as used in this workshop, is defined as the process of determining the value or effectiveness of some activity for the purpose of decision making. Decision makers need information to determine whether to continue, modify, expand, or discontinue individual programs. Program Evaluation is seen as the source of that information.

Avoid drawn-out technical discussion and argument. Accept the fact that other definitions may be valid.

2. Ask the participants to review the reading on definition starting on A-5.

3. TR. A-3 State that the three major phases of evaluation are Plan-Conduct-Use (A-6 of the Guide). Suggest that typically more time is spent on conducting than planning or using. Hopefully this workshop will demonstrate the crucial need for planning and using as well as conducting.

4. TR. A-4 Identify the eight steps of the evaluation process (A-6 of the Guide) and put this section on Purposes and Requirements in perspective to the total program evaluation and the program management activities.

5. TR. A-5 Briefly discuss the different types of evaluation data which may be collected and the different ways they may be utilized in the evaluation process. In presenting the transparency (A-7), cover the Product, Process and Context portion so that only Formative and Summative are viewed. At this time distinguish between formative and summative evaluation. Formative evaluation is concerned with the implementation of the program and interim information on the students' progress toward attainment of stated objectives. Formative evaluation takes place during a program. Summative evaluation looks at the final outcome--whether or not stated objectives were reached--and takes place at the end of a program.

   Uncover the Product, Process and Content section. Differentiate between the three types of data that should be collected in an effective program evaluation.

   - Product refers to the program outcomes. Learner achievements of one kind or another are the expected outcome of every instructional program. There could be exceptions such as in a program designed to produce materials, but the emphasis in this workshop is on the evaluation of instruction.

- Process refers to the activities used to support the achievement of expected outcomes. This may include data such as instructional processes and inservice training activities.
- Context refers to the environment in which the activities take place. It may include factors such as facilities, scheduling, rules and regulations, and sociological climate.

6. Ask the participants to review the discussion on Types of Evaluation Data on pages A-7 and A-8. Call attention to the examples given.

7. Introduce Learning Exercise 1, Types of Evaluation Data, on pages A-25 and 26 of the Guide. Ask each group of participants to arrive at consensus on the four items that illustrate product, process, or context. Compare answers with those on the sheet on page A-26. Allow time for discussion.

8. TR. A-6 Discuss briefly the cycle of program management and evaluation, the feedback from one step to another, the mutual dependency of all of the steps, and the centrality of evaluation in the process (A-9).

9. Ask the participants to review the chart and discussion on pages A-9 and 10. Answer questions and discuss further if required.

Role of the Evaluator

1. We have discussed the purposes of evaluation, two types of evaluation processes, and the kinds of evaluation data to be collected. The next question to which we must address ourselves is "Who will be the person responsible for the evaluation?"

2. Questions to further stimulate a discussion on the role of the program evaluator might be as follows:

- Should the program evaluator be from within the district or external to it?
- What sources exist - both internally and externally - for assistance in the conduct of program evaluation?
- What are the advantages of an internal program evaluator? Disadvantages?
- What are the advantages of an external evaluator? Disadvantages?
- What professional and personal attributes should the program evaluator possess?

3. Call attention to the short reading on Role of the Program Evaluator on pages A-11 and 12 of the Guide.

Initial Steps of the Evaluator

1. After the evaluator has been selected, he will need to address himself to certain questions before the development of evaluation standards.

2. TR. A-7 Project the transparency and briefly discuss each of the questions (A-13).

3. Suggest that one of the next steps of the program evaluator is to take a close look at the needs statements, the goals and the objectives of the program.

4. Talk through or ask participants to read pages A-13, 14, 15 in the Guide (five minutes). Review the main points again emphasizing the importance of the evaluator's examining the needs assessment, program goals and objectives, and judging how well they relate to the program activities.

   Follow-up questions to the reading might include:

   - What should the evaluator look for when he or she reviews an existing program?
   - Why is it important to know whether and how needs were identified?
   - What might the evaluator seek to learn when reviewing program goals and objectives prior to beginning with the evaluation planning?

Questions about needs-assessment techniques may arise. The topic is not discussed at length in this workshop; however, a good reference is An Analysis of Needs Assessment Techniques for Educational Planning at State, Intermediate, and District Levels written by Belle Ruth Witkin, Ph.D. and published by the Alameda County School Department in Hayward, California. It is available through ERIC (identification number: ED 108370).

5. TR. A-8 Distinguish between goals and objectives. (See A-15.)

6. TR. A-9 Comment briefly on each of the criteria for working performance objectives. (See A-15 and 16.)

7. TR. A-10 Ask the group to identify the six components of the performance objectives shown on the transparency.

8. Introduce Learning Exercise 2, Identification of Measurable Objectives, on pages A-27 and 28 of the Guide. Ask table groups to arrive at a consensus on each item. Review and discuss with the total group when subgroups have finished.

   Option: Ask each small group to select one of the objectives in the exercise which does not satisfy the requirements of measurability and rewrite it according to the six criteria.

9. Introduce Learning Exercise 3, Selecting Appropriate Objectives, on pages A-29 - 38 of the Guide. Ask the participants to arrange themselves in groups of five or six. Each group should choose one of the four goals and match a number of appropriate objectives to it. Review all materials in the exercise before asking the participants to begin.

   TR. A-11 Distribute to each group a response transparency identical to
   TR. A-12 the Group Report Form on page A-37 so that the group ratings
   TR. A-13 may be recorded. Ask a leader from each group to display its
   TR. A-14 transparency to the entire group.
   TR. A-15

TR. A-16 After each report, flash the corresponding comparison transparency
TR. A-17 for self-correcting feedback. Review the objectives that go
TR. A-18 with the numbers shown.
TR. A-19

10. Discuss the need for a review of program activities to insure their appropriateness for stated program objectives. (See Summary paragraph at page A-16.)

11. Introduce Learning Exercise 4, Matching Needs Statement to Program Objectives and Program Activities, at page A-39 of the Guide. Ask each small group to match the needs statements with appropriate program objectives and activities by writing the number of the related needs statement next to the matching program objective and program activity. Review the results with the total group.

12. TR. A-20 Review topics covered to this point in order to maintain the perspective of the total workshop.

Requirements of Program Evaluation (at page A-16 of the Guide)

1. Different decision makers--teachers, principals, superintendents--have different evaluation-information requirements,and the questions of each must be considered in the early stages of evaluation planning.

2. TR. A-21 Project the transparency and discuss three key questions to be considered when determining the requirements of the particular program evaluation (A-16 - 17). Ask if there are other questions that should be considered.

3. Talk through or ask the participants to review the reading on Requirements of Program Evaluation on pages A-16 and 17 as a review of the foregoing discussion.

4. Review with the participants the suggested Evaluation Information Requirements form on pages A-18 and 19 of the Guide.

5. Ask the participants to review the short reading on End of Year Evaluation and Interim Evaluation on A-20. Possible questions to ask after the reading has been completed are as follows:

   - How do interim evaluation requirements differ from end-of-the-year requirements?
   - Should the end-of-the-year evaluation include interim evaluation data? Why or why not?

6. Next, review the part on Identifying Resources and Constraints on A-20 and 21. Ask:

   - What evaluation resources are likely to be available to a program? What are some likely constraints?

   Recall a constraint in a particular program that was satisfactorily resolved.

7. Review with participants the summary of the tasks performed by a program evaluator at the purpose and requirements stage on A-22.

8. Call attention to the Checklist of the Steps in Determining Evaluation Purposes and Requirements on A-23 and 24.

9. Present the Purpose section of the Workbook on Program Evaluation at pages 1-4. Emphasize its programmed instruction approach.

10. TR. A-22 Project this transparency as a bridge to the next section on Planning.

EVALUATION TRAINER'S GUIDE

Section B

PLANNING

The Evaluation Improvement Program

TRANSPARENCIES

Transparencies	Subjects	Trainer's Guide	Evaluator's Guide
TR. B-1	The Evaluation Process	B-1	B-1
TR. B-2	First Four Planning Steps	B-2	B-1,2,3
TR. B-3	Five Later Planning Steps	B-2	B-2,4,5,6
TR. B-4	Review Evaluation Plan	B-5	B-6
TR. B-5	Planning Saves Time and Energy	B-6	B-9

LEARNING EXERCISE

5	Planning for Assessment of Program Objectives	B-11

Introduction

The purpose of this section is to present an outline for a general program evaluation plan. Emphasis is placed on the need for careful planning prior to the implementation of the evaluation process and as the program to be evaluated is being planned.

1. TR. B-1 Project transparency on screen while giving an introduction to the section.

> In the last section, we discussed the different purposes and requirements of evaluation as perceived by various audiences. Then we took a look at the initial steps the program evaluator takes as he begins to develop a plan. Now we shall look at the sequence of steps essential to developing a total program evaluation plan. You will note that we are at Section B, Develop an Evaluation Plan.

2. > In the development of an evaluation plan, there are a number of steps that should be included. While there is no universal list or sequence that is best for every situation, the steps we have developed should be of assistance as you develop a plan for program evaluation. Let's briefly review the steps.

Suggestions for Teaching

The trainer should have become well acquainted with the discussion of steps of evaluation planning on pages B-1 through 6 of the Guide. During the presentation, this discussion should be referred to and each step expanded upon through classroom interaction. It is important, however, that the review of the steps be limited to no more than 10 minutes. If detailed questions are asked, emphasize that this is an overview and further discussion of each step will come later.

TR. B-2

- Review the statements of need
- Match the program goals and objectives with the statements of need
- Determine the questions for which the evaluation design should provide relevant information — B-1,3
- Identify the general type of assessment instruments which will provide the most appropriate kinds of information

TR. B-3

- Determine in general terms the most appropriate time for securing the needed information and who should do it
- Select the kinds of data analysis procedures which will be most meaningful for the data collected — B-4,5,6
- List the activities to be monitored
- Identify the deadline dates for all required reports; identify all persons or agencies to receive reports
- Determine use to be made of each report or kind of data reported

Program Evaluation Planning Form

1. Turn to page B-2 of your Guide. There you will find the planning steps in chart form. This chart can be used by the program evaluator in developing a general evaluation plan.

2. Introduce Learning Exercise 5, Planning for Assessment of Program Objectives, on pages B-11 through 15 of the Guide. Ask the participants to work in small groups, with each group devoting itself to one of the four objectives selected by Rosedale for its upcoming program evaluation.

3. Ask each group to agree and record the title of the program chosen, apparent purposes, and a key audience in the appropriate spaces on page B-14.

4. Briefly discuss the questions in the first two rows of the Program Evaluation Planning Form on B-15:

   Program Objectives
   Evaluation Design

   and then ask the groups to apply the questions in these categories to their selected programs and record their responses in the appropriate spaces on the work sheet on B-14.

5. Next, discuss the questions in the next two rows:

   Assessment Instruments
   Administration Dates and Personnel

   After the brief discussion, ask the groups to complete the appropriate rows. Responses should be the result of consensus among members of the group.

6. Briefly discuss the questions in the Data-Analysis Techniques row and then ask the groups to complete that row on their work sheet.

   NOTE: Advise the participants that they should not be concerned if they are unable to respond to all of the questions. By the end of the workshop, they will have sufficient information to do so.

7. Briefly discuss the questions in the next two rows:

   Monitoring Program Activities
   Monitoring Dates and Personnel

   and ask the groups to complete these rows.

8. Discuss the questions posed in the final rows:

   Key Reporting Dates
   Personnel or Agencies to Receive Reports
   Use to be Made of Data or Reports

   and ask the groups to complete the work sheet.

As the groups are working on the questions in each row of the work sheet, the trainer should circulate among the groups and assist as necessary.

9. As a culmination activity, ask a spokesperson from each of several groups to report in turn on their results. The trainer should be alert to inappropriate plans that might be reported and indicate that more information in this area will be given later in the workshop.

Review of Preliminary Plans (at B-6)

1. TR. B-4 We now have a preliminary program evaluation plan. The next step is to have the plan reviewed by all of those persons who are involved in the program or who may be interested in its evaluation. Staff members will be far more supportive of a program evaluation if they feel they had some input during its planning. Errors or misconceptions that may have gotten into the plan may be corrected. And, too, the identification of potential problems in the planning stage may avoid difficulty during execution or opposition at the time of reporting results.

Evaluation Timeline (at B-7 and 8)

Now let's take a look at planning the management aspects of a program evaluation.

1. A timeline is developed for four purposes:

- To use as a communication tool
- To gain support and commitment among participants
- To schedule designated activities
- To monitor scheduled activities

As a communication tool, the evaluation timeline becomes the calendar for all major activities. School personnel involved in or affected by evaluation activities need to be notified of forthcoming events as well as being included in the planning and scheduling of some of the evaluation activities. As a tool for helping to get the evaluation done, the timeline provides a concise picture for the participants and manager of evaluation activities.

The timeline should show each major activity, the time-frame for accomplishing the activity, and the personnel involved in completion of the activity. The timeline also becomes a useful tool when evaluation events go off schedule. Revision dates can be added and the effects of the revision on other events can be viewed.

It is important to remember that timelines and activities generally go through revision throughout the year. Don't discard your timeline when events or dates begin to change. Be flexible and change those sections of the timeline that have undergone revision, holding to the basic pattern.

2. Review the timeline form on page B-8 of the Guide. State that this is only one example and ask the participants to describe timeline charts with which they have worked.

   - With whom should timeline charts be shared during implementation of the program evaluation?

Determining and Obtaining Required Resources (at B-7 and 9)

1. The next step for the program evaluator is to ask:

   - What resources are required for implementing the program evaluation planned?
   - What resources are available?
   - How can descrepancies between resources required and resources available be rectified?

2. Talk through or ask the participants to read Determining and Obtaining Required Resources on pages B-7 and 9 of the Guide.

Checklist of Major Steps

1. Call attention to the Checklist of the Major Steps Required in Developing a General Evaluation Plan on page B-10 of the Guide.

2. Present the Needs Assessment, Goals and Objectives, and Data Requirements sections of the Workbook on Program Evaluation (pages 5-24). If time permits, go through the first steps of working with one of them.

> TR. B-5 This is the completion of the section devoted to developing a general program evaluation plan. Planning your evaluation carefully will not only save you time but also energy and will relieve you of any anxiety you may be feeling.
>
> Our next action will be devoted to determining the appropriate evaluation design for addressing the questions that need to be answered about the program being evaluated.

EVALUATION TRAINER'S GUIDE

Section C

DESIGN AND SAMPLING

The Evaluation Improvement Program

TRANSPARENCIES

Transparencies	Subjects	Trainer's Guide	Evaluator's Guide
TR. C-0	The Evaluation Process	C-3	C-1
TR, C-1			
TR, C-2	Are the Students Entering this Program Really Functioning at Level Expected?	C-3	C-2
TR. C-3	At What Level Are Students Functioning at the End of the Program?	C-3	C-2
TR, C-4	How Much Did Students Improve During the Program?	C-3	C-3
TR, C-5	Before/Intervention/After	C-3	C-4
TR, C-6	How Much Did Students Improve During Program?	C-3	C-4
TR, C-7	How Do Nonprogram Students Identical with Program (1) Students Compare with Regard to Growth?	C-3	C-6
TR, C-8	How Do Nonprogram Students Identical with Program (2) Students Compare with Regard to Growth?	C-3	C-6
TR, C-9	How Do Comparable Nonprogram Students Compare with Program Students?	C-3	C-7
TR, C-10	How Do Program Students Compare with Some Norm Group?	C-3	C-7
TR, C-11	How Do Students' Achievements in the New Program Compare with Their Past Performance?	C-3	C-8
TR, C-12	Random Sample	C-7	C-20 - 23
TR. C-13	Table of Random Numbers	C-8	C-26
TR. C-14	Number of Students	C-8	C-28
TR. C-15	Students for Classroom Observations	C-8	C-31

LEARNING EXERCISES

6	Evaluation Design	C-16 - 19
7	Sampling Considerations	C-24 and 25
8	Random Sampling	C-28 - 31

Introduction

The purpose of this section is to present a useful but nontechnical discussion of evaluation design. The background given is sufficient to the task of selecting a design in a given set of circumstances. The section also discusses sampling and emphasizes situations in which the use of sampling procedures is appropriate.

Topics Covered in This Section

- How to Select a Design
- The Level at Which Students Are Functioning
- Gains Made During the Program
- Gains Made in Comparison to Expectations
- Elements of the Program Contributing to Gains
- Interpretability of Results
- Pitfalls to Avoid
- Factors to be Considered
- Monitoring Activities
- When to Use a Sample
- How to Select a Random Sample
- Stratified Random Samples
- Sample Size
- Case Studies

Things to Stress in This Section

1. The basic question in program evaluation is whether the program resulted in improved learning outcomes over previous or current programs. (This point should be periodically reinforced.)
2. The program evaluator should work with the program manager while the program is in the planning stage.

3. Design begins when you know what questions will have to be answered at the end of the evaluation.

4. The better designs involve some kind of reference group or reference data.

5. Random assignment of individual pupils to different programs will do more to assure that groups are alike than using intact classes.

6. Factors which may influence the program evaluation results should be anticipated in advance and taken into account in choice of evaluation designs.

7. The evaluation design must be chosen as soon as feasible--ideally when the program itself is planned, hopefully before the program starts, but certainly well before data collection has begun.

The question of primary importance in program evaluation is: How much more did pupils learn by participating in the program than they would have learned without it?

To find out whether students have done better in a special program than they would have done without it requires information about:

1. How much students improved
2. An accurate estimate of how they would have done without the program

Evaluation Designs

Suggestions for Teaching

1. TR. C-0 Tie evaluation design to other parts of the evaluation process.

2. Present the designs given on transparencies C-1 through C-11 and refer to the appropriate pages of the Guide as the presentation proceeds.

   TR. C-1 Warm Feeling--Stress the lack of objectivity in this type of evaluation and that this does not constitute good program evaluation (p. C-1).

   TR. C-2, TR. C-3 Indicate that these do not constitute good program evaluation either (p. C-2).

   TR. C-4, TR. C-5 Communicate the idea that these designs are better in terms of giving more information but they still do not answer the questions about how worthwhile this program is (pages C-3 & 4).

   TR. C-6, TR. C-7, TR. C-8, TR. C-9, TR. C-10, TR. C-11 Of these alternative designs, stress that random assignment is to be preferred if at all possible (p. C-5 through 8).

Elements of the Program Contributing to Gains

What to Stress (at C-8-11)

Indicate that factorial designs such as these give added information if they can be implemented successfully. The unit of analysis can be either entire classrooms or individual students. Whatever units are selected, they must be randomly assigned to the various cells in the table. If entire classrooms are used as units and there are two classrooms per cell, then N=2 (rather than the total number of students in the two classrooms). Students are counted only if they have been randomly assigned on an individual basis to the various conditions.

Suggestions for Teaching

1. Summarize the points given in the opening paragraphs on C-8.
2. Ask participants to turn to page C-9 and look at the two factorial designs as you describe them.
3. Have participants turn to the design on page C-11 and explain the difference in units of analysis (i.e., classrooms vs. students).

Interpretability of Results

Suggestions for Teaching (at C-12)

This discussion and the one on Pitfalls to Avoid are important to consider if designs are to be properly carried out and misinterpretations avoided.

1. Discuss each of the points mentioned on C-12. This material can be used to generate interesting discussions since participants often will have had similar experiences that have tended to invalidate evaluation results. The point is that the program evaluator must be alert to extraneous problems which can lead to erroneous conclusions about the effectiveness of a program.

   There are three sets of instructions under Pitfalls about which questions often arise. These are discussed in 5,6, and 7, on C-14. If anyone questions these instructions, here are the answers:

   (5) The reason you should not use a selection test as a pretest is a statistical artifact. Because of error inherent in all measurement, extreme groups will always contain some students who have performed atypically well or atypically poorly. If you gave the group the same test the next day with no intervening instruction, some pupils would score much higher (or lower) than they did on the previous day. This is called "regression toward the mean." If different tests are not used, results from pretest-posttest comparisons will be contaminated.

(6) There are two primary problems with grade equivalents: 1. Grade equivalent "scores" do not qualify as score data; they do not meet the assumptions for t-tests, analysis of variance, etc. 2. Most normative data are collected at one or two specific times during the year. At other times during the school year, values are inferred by fitting a curve to the actual data from those collections. This assumes that learning proceeds uniformly over the school year. There is substantial research evidence that says this just is not so.

(7) Computing gain scores, the difference between pre- and posttest on a single group,has statistical limitations due to differential growth rates, different test score reliabilities as a function of achievement level, and test floor effects. These factors work to invalidate results when inferential tests are used. There are two kinds of gain scores: raw gain scores and residual gain scores. A raw gain score is the difference between pre- and posttest. A residual score is the difference between a posttest score and some estimated or expected posttest score, where the expectation has been derived from the pretest scores using regression techniques. Raw gain scores systematically overcorrect the differences between program group and reference group; residual gain scores systematically undercorrect.

2. Stress the kinds of factors which must be considered in planning a sound design. (See C-14 and 15.) This section can be used to generate a discussion if the trainer wishes. Most participants can add to the list of factors if given the opportunity.

3. Finally, stress the need to monitor program activities in order to assure implemetation according to plan. (See C-15.) Not all goes according to plan unless someone has the specific job to see that it gets done.

4. Proceed with Learning Exercise 6, Evaluation Design, at pages C-16 through 19. There are good designs other than the solution presented. For example, a factorial design would be superior if it could be implemented.

5. Summarize the main points on evaluation design before proceeding to sampling.

Sampling

Things to Stress (at C-20)

1. Past experience indicates that some confusion exists about when to sample and when not to sample. This issue is addressed in your opening introductory statement on the following page and in the discussion (How Large Should a Sample Be?) on page C-9 of this Guide. Periodically reinforce the idea that sampling for program evaluation purposes should be used primarily with large settings (i.e., parents, community members, large school districts). Sampling within school programs is not usually feasible for program evaluation purposes unless a factorial design is used or the number of students in the population of interest is large. Within-classroom sampling may be useful for certain purposes.

2. Sampling is a good way to get more information at less cost about a total program.

3. Using a table of random numbers is a simple procedure and the soundest one to use when drawing a sample.

4. A stratified random sample gives greater assurance that the group studied will be representative of the total population.

When information is needed on large numbers of persons, it is always more economical to sample a portion of the total number on whom you would like to get the information. If sampling is properly done, you can infer what the results would have been, had it been possible to obtain data on everyone. The decision to sample for program evaluation must be made in relation to the size of the group on whom you want information. Sampling within classrooms or programs having relatively few students is not defensible in serving program evaluation purposes though it may be useful in providing information that will help the teacher make decisions about day-to-day teaching plans. If there are large numbers of classrooms involved in a given program, then a sampling of students from each classroom may well be worth considering. The sampling of parents and community groups, however, is a practice almost all schools and districts will want to consider.

Suggestions for Teaching

1. After the introduction above, the Introduction to Sampling at C-20-23 in the Guide may either be delivered in lecture form or talked through, or the participants may be asked to read it themselves.
2. Introduce Learning Exercise 7, Sampling Considerations, on pages C-24 and 25 by reading the situation. Give participants the option of working alone or in groups. When most participants seem to have finished, review items presented on comparison sheet and ask if anyone has additional points to make.

How to Select a Random Sample (at C-26)

TR. C-12 The method of selecting a random sample is not complicated, but there are some common misconceptions as to how a random sample is drawn. The random selection of a sample must give each person in the population an equal chance to be drawn in the sample.

Suggestions for Teaching (at C-26)

1. Following the introduction, demonstrate use of Table of Random Numbers as on C-26 and 27.

   TR.C-13 Table of Random Numbers. Use to demonstrate how numbers are selected.

2. Introduce Learning Exercise 8, Random Sampling, at pages C-28 through 31, and give participants 5 to 10 minutes to complete it. They may work together if they wish to do so. As they complete the task, have them compare their selections to those given on the comparison sheet. Help clarify misunderstandings.

   TR.C-14

   TR.C-15

> Another way to handle the problem is to draw a stratified random sample. In a stratified random sample, the population is first divided into categories or strata and then random samples are selected for each category or stratum. The more catagories that you include, the less you have to depend on randomization to handle the extraneous or uncontrolled factors. This is so since the units within a randomly sampled stratum will be alike on the category selected for stratifying.

Suggestions for Teaching (at C-32)

1. Example: Ask participants to turn to page C-32 in their Guide. Work through the successive divisions and subdivisions to the final stratification by sex, number of periods, and grade level on C-33.
2. Talk through the reprise at the end of the example. As an optional discussion, you might ask how the group would stratify a community if they were to do a survey. Possible stratification variables might be:

   a. occupational level of parent or other participant
   b. children in school vs. children not in school
   c. bilingual vs. monolingual

3. Example: Similarly, on C-33 and 34, work through the stratified random selection of test items. Be sure to note that the number of items in the cell do not have to be equal.
4. End this discussion with a brief mention of cluster sampling as at the bottom of C-34.

How Large Should a Sample Be? (at C-35)

> Sampling students or student groupings within large schools or districts for program evaluation purposes is practical, but for smaller schools and districts it is not. However, sampling of parent or community groups is practical for all except the smallest of school settings. The crucial factors are the size of the population and the amount of error that population size imposes. The amount of error the evaluator is willing to tolerate is what determines the practicality of using a sample. "Population" in this sense refers to the group for whom you want information--it may be all fifth graders, or all parents of secondary students enrolled in noncollege preparatory curricula, or all adults of voting age in the community.

Suggestions for Teaching

1. Present the example in the text at pages C-35 and the Table for Determining Sample Size at C-36.

Two Case Studies on Sampling (at C-37)

Suggestions for Teaching

1. Center a discussion on one of the two Case Studies at pages C-37 through 40.
2. If time is short, refer to the other in passing for later review.

EVALUATION TRAINER'S GUIDE

Section D

INSTRUMENTS

The Evaluation Improvement Program

TRANSPARENCIES

Transparencies	Subjects	Trainer's Guide	Evaluator's Guide
TR. D-1	The Evaluation Process	D-4	D-13
TR. D-2	General Characteristics of Standardized Instruments	D-4	D-13
TR. D-3	Reliable But Not Valid	D-4	D-14
TR. D-4	Neither Reliable nor Valid	D-4	D-15
TR. D-5	Reliable and Valid	D-4	D-15 and 16
TR. D-6	What is Being Measured?	D-4	D-20
TR. D-7	What Kinds of Instruments Will Be Used?	D-5	D-20
TR. D-8	What Kinds of Item Types Are There?	D-5	D-21
TR. D-9	What Kinds of Scores Are There?	D-5	D-21
TR. D-10	Raw Score, Grade Equivalent	D-5	D-23
TR. D-11	Major Uses of the Interview Technique	D-7	D-28
TR. D-12	Likert-Type Item from a Teacher Questionnaire	D-7	D-36
TR. D-13	Summary of Questionnaire Item Structure	D-7	D-42
TR. D-14	Observation Record	D-8	D-53
TR. D-15	Assuring Rater Reliability on Observation Instruments	D-8	D-57
TR. D-16	Objective 1 Bilingual-Bicultural Understanding	D-10	D-69
TR. D-17	Objective 2 Reading Comprehension	D-10	D-69
TR. D-18	Objective 3 Language Usage	D-10	D-69
TR. D-19	Objective 4 Remedial Mathematics	D-10	D-69
TR. D-20	Instrument Development	D-11	D-78
TR. D-21	Summary of Basic Evaluation Instruments	D-11	D-80

LEARNING EXERCISES

9	Reliability and Validity	D-17 - 19
10	Judging Items	
11	Criticizing a Classroom Observation Instrument	D-59 - 62
12	Selecting Norm-Referenced Tests	D-69 - 73
13	Selecting Appropriate Instruments	D-74 - 75

Introduction

The central purpose of this section is to help make workshop participants aware of where to turn for help in identifying and selecting instruments to use and when to develop needed instruments locally. Beyond that, there is an orientation to educational measurement and a survey of each of several types of instruments.

Topics Covered in This Section

- Scenarios to Set the Stage
- Standardization
- Reliability and Validity
- Norm-referenced Tests
- Criterion-referenced Tests
- Questionnaires
- Observational Techniques and Instruments
- Other Behaviors
- Where to Find Existing Instruments
- The Development of Instruments

Things to Stress in This Section

1. Program evaluation needs to be planned in order to prevent the kinds of things that happen in the scenarios.
2. The careful selection of program and program evaluation objectives and of the evaluation design must precede determinations regarding assessment instruments.

Suggestions for Teaching

SCENARIOS

Now that you have seen the kinds of questions that must be addressed in the planning stages, we are going to take time, through a set of three scenarios, to show what often happens in school districts. Copies of the scenarios are in your manuals, pages D-1 through D-8. They illustrate some important general principles of evaluation which we will discuss afterward.

The first scenario is entitled The Report: Useful to Whom and for What? Suppose that you are the program evaluator of a reading program, waiting to see the principal (identify principal) of the school who is busy talking to Mrs. Smith (indicate Mrs. Smith). You know Mrs. Smith. She is a reading teacher and you have been in her classroom once or twice. Your impression is that she is a very outspoken type of person and that she often talks a great deal.

The door of the principal's office is not closed and you hear the following conversation.

1. At least a half-day before this point is reached in the workshop, announce the three skits, describe the characters and select persons to play each part. The assignment of persons to specific roles is crucial to the success of the scenarios. Previous experience has shown the skits to be highly successful or very dull, depending upon the spirit of the cast. Not everyone enjoys role playing. Select the most out-going persons for the major parts (Mrs. Smith, the Principal, Chairman, Mr. Worth, Mr. Fairchild). If you have worked with the group previously, you will know who will do the best job. If not, check with the trainer who has worked with the group on previous sections in order to line up good prospects for the parts.

2. Discussion can follow each scenario or be held after all three have been done. At any rate, this material is generally a good "warm up," and it is fairly easy to encourage group discussion about the various messages contained in the scenarios. A list of points illustrated in the scenarios follows the last one, at page D-9 in the Guide. It should generally be used as a check to see if the participants have spontaneously recognized the significant issues.

Overview and Basic Considerations (at D-11 of the Guide)

Things to Stress in This Section

1. Periodically mention the desirability of using multiple measures for a given program objective.
2. The program evaluator should spend time searching in the library, familiarizing himself with local resources.
3. Note the broad definition of what standardization is. There are standardized questionnaires and observation records as well as achievement tests.

The remainder of the workshop is designed to help you avoid the pitfalls illustrated in the scenarios. We have already talked about determining purposes and requirements for program evaluation as well as the development of an overall plan and the program evaluation design you might plan to use. This section will concentrate on the instruments used to obtain data on program effectiveness.

Suggestions for Teaching

1. Begin with introductory transparancies.

   TR D-1 Tie instrument selection to the other elements in program evaluation.

   TR D-2 Read the four points on Characteristics of Standardization (page D-13).

   TR D-3, TR D-4, TR D-5 For definitions and discussion of reliability and validity (pages D-13 through 16).

2. Direct participants to Guidelines on Reliability and Validity Coefficients on page D-16.

3. Introduce Learning Exercise 9, Reliability and Validity, at D-17 through 19 of the Guide.

Some participants have trouble trying to handle the nondefinitive answers given to this exercise. Actually, the point of this exercise is that not all is cut and dried or black and white in instrument selection. You cannot always blindly follow a set of rules. There is a point where you must use your own judgment and common sense. In some cases, more information is needed before a rational judgment can be made. It is important for the program evaluator to recognize this and to try to find the additional information needed.

Types of Assessment Instruments (at D-20 of the Guide)

Achievement Tests

TR. D-6 The selection of assessment instruments begins with the general questions of what is to be measured. A somewhat arbitrary, but convenient, classification is presented here. The first step is to think about what it is you want to measure.

- achievement
- attitudes
- interaction among persons
- other behaviors

Suggestions for Teaching

1. Then follow in sequence with:
   TR.D-7 Kinds of instruments
   TR.D-8 Kinds of items
   TR.D-9 Kinds of scores

2. TR. D-10 Briefly review the table at D-25 on Types of Test Scores.

3. Compare norm-referenced tests with criterion-referenced tests (D-22 and 23). "Criterion-referenced" has become a catchall term, not only for measures appropriately described by it, but also for measures perhaps more precisely defined as content-(or objectives- or domain-) referenced or as construct-referenced.

   It is probably not necessary and may be undesirable to try to make these distinctions with a workshop group. However, should participants inquire, you might use these examples:

   a. "Criterion-referenced" means there is a definite pass-fail point such as in the setting of mastery scores.

   b. With content- (or objectives- or domain-) referenced measures, no such external pass-fail requirements are set, but agreement has been reached on objectives,in a given domain, and the content has been adequately defined. Scores on the instrument have meaning in their own right without reference to external criteria or norms. The National Assessment of Educational Progress typifies this type of measurement.

   c. Construct-referenced measurement is used when "it is important to characterize individuals in terms of fundamental traits (e.g., intelligence, neuroticism) or to characterize situations in terms of fundamental properties (e.g., liberal, totalitarian)."*

---

*Anderson, S.B., Ball, S., & Murphy, R.T. Encyclopedia of educational evaluation. San Francisco: Jossey-Bass, 1973.

4. Emphasize the point that a norm-referenced grade equivalent score is a measure of central tendency.

5. Summarize the common types of scores found on norm-referenced tests (D-21 and 25). Spend sufficient time on discussion of the table on D-25. An understanding of properties of different kinds of scores is essential to the selection of appropriate descriptive statistics and inferential tests. Typically, the raw score is the best one to consider when using inferential tests for score data. Neither grade equivalents nor percentiles are score data. They are ordered data.

6. Go on to criterion-referenced tests and draw the distinction between class mastery (D-23) and student mastery (D-24).

7. Review some sources where criterion-referenced tests may be found (e.g., table on D-26).

Questionnaires (starting at D-27)

Things to Stress in This Section

1. Look for existing instruments before developing new ones.

2. Decisions about how the questionnaire is used to collect data must be carefully made. Higher response rates can be achieved if instruments are administered in a captive group. But this may not give an unbiased sample of responses. Mailed questionnaires are easy to distribute but difficult to get returned. Special populations or unusual problems may necessitate one-to-one interviews.

When attitudes are to be measured, frequently some kind of questionnaire is used. There is a wide variety of questionnaire item types and techniques from which the program evaluator may choose. Knowing the range of possibilities and the advantages and disadvantages of each makes the task easier for the evaluator.

Suggestions for Teaching

1. Introduce the questionnaire as a widely used means of collecting various types of survey information. Distinguish between surveys used in a group-response situation or sent home with students or mailed and those used as a guide to conduct an interview.

2. TR. D-11 Summarize where interviews are most useful.

3. Have participants read the section on Guidelines on Review and Selection on pages D-28 and 29.

4. Discuss open-ended items and how they are categorized (D-30 through 34).

5. Have participants turn to page D-34 and summarize the advantages and disadvantages of open-ended items in questionnaires.

6. Have participants follow in their Guide as you present the various kinds of objective items for use in questionnaires (D-34 through 41).

7. Go into some detail on rating scales (D-36 and 37).
TR. D-12 This is an example of Likert Scales. Stress the need to define scale points.

8. Introduce the Kropp-Verner Scale (D-37 and 38). This is an example of a Thurstone Scale in which a series of value statements have been empirically assigned numerical weights that correspond to their judged worth or value. On page D-39, which illustrates the scoring of the Kropp-Verner Scale, tallies in Column 2 are shown as frequencies in Column 3. Column 4 gives the numerical value assigned to each statement on the basis of previously judged value in the process of standardization. This value can be used to obtain individual scores on persons by adding the value for each statement checked and then computing the median. However, group interpretation can be made simply by computing the median of the total group frequencies.

9. Discuss how rankings differ from ratings. Refer to these two variations of the same item (D-40 and 41). Review with the group or have them do so individually.

10. TR. D-13 Summarize the item structures for questionnaires (D-42).

11. Introduce Learning Exercise 10, Judging Items, at D-43, and let participants work in small groups. After about 10 minutes, review the results.

Observational Techniques and Instruments (at D-48)

Things to Stress in This Section

1. If observational techniques are to be used, make sure they are sufficiently structured and that persons who use them have been sufficiently trained. Never use an observational method in program evaluation unless you have first checked the interrater reliability.

> While both achievement tests and questionnaires can give valuable information for program evaluation, there are certain kinds of needed information that cannot be obtained from them. Observational techniques and instruments for recording observations provide an added dimension. As with any other type of assessment instrument, there are both advantages and disadvantages in using them.

Suggestions for Teaching

1. Ask participants to turn to page D-48 and follow table giving advantages and disadvantages of observational techniques.
2. Introduce Example 1 on how to observe question-answer-feedback sequences in classrooms (D-49 through 51).
3. TR.D-14 Introduce Example 2 on measuring interaction among students (D-51 through 54).
4. Refer briefly to D-54 through 56 on the amount of detail required for reliable use of a system such as this.
5. TR.D-15 Explain how to assure interrater reliability (D-57).
6. Summarize steps in selection of an observation instrument (D-57 and 58).
7. Introduce Learning Exercise 11, Criticizing a Classroom Observation Instrument, at D-59 through 62. Have participants work in small groups. Discuss the results after 5 to 10 minutes.

Other Types of Instruments (D-63)

The last type of instrument to be discussed is the unobtrusive measure. The greatest advantage in using this type of measure is that data collection procedures do not influence the results. Data are collected without knowledge of the individuals involved and sometimes can be drawn from existing information files. Such things as attendance records; notations of awards, citations, and honors; conventional library records; histories of dropouts and transiency provide raw data for unobtrusive measures.

Suggestions for Teaching

1. Read a few indicators from the listings at D-64, D-65, and D-66 to illustrate a range of unobtrusive measures.

Sources of Information about Instruments and the Selection Process (D-66)

Buros' Seventh Mental Measurements Yearbook is the best known source of published assessment instruments, but it is now somewhat dated. The references listed here give fairly comprehensive coverage over a wide range of instrument types.

Suggestions for Teaching

1. Have participants turn to page D-67 and look at the table that shows where to locate information about tests. Discuss each briefly and show examples if you have them.
2. Ask participants to turn to page D-68 and follow Factors to Consider in Selecting Instruments while you briefly summarize them.

3. Introduce Learning Exercise 12, Selecting Norm-Referenced Tests (See D-69 through 73). After a short period of work, go over selections.

	Objectives	Instruments
TR. D-16	1	None appropriate. Instruments to be developed.
TR. D-17	2	Cooperative Test of Basic Skills Cooperative Primary Tests Gates-MacGinitie Reading Tests Metropolitan Achievement Tests
TR. D-18	3	Circus Test of Basic Experiences Stanford Early School Achievement Test
TR. D-19	4	Comprehensive Tests of Basic Skills Metropolitan Achievement Tests

Using Existing Instruments or Using Your Own (D-76)

Every type of instrument that has been discussed in this section is accessible somewhere. The program evaluator who spends time searching for instruments that will meet his program needs will be far ahead of the one who decides to launch a school-wide or district-wide effort to develop locally made criterion-referenced tests, questionnaires, or observation records.

Suggestions for Teaching

1. Ask participants to read pages D-76 through D-79.
2. TR. D-20 When participants have finished, discuss steps in instrument development (D-78).
3. Comment on need to establish reliability and validity of locally developed instruments.

Section Review

Suggestions for Teaching

1. Summarize the seven review points on D-79.
2. Introduce Learning Exercise 13, Selecting Appropriate Instruments (see D-74 and 75). Ask participants to work in small groups. Discuss exercise after a short period.
3. Present Instruments section of Workbook on Program Evaluation, pages 35-38. Emphasize importance of matching instrument to objectives to be measured.
4. TR. D-21 Summary of evaluation instruments (D-80).

EVALUATION TRAINER'S GUIDE

Section E

DATA COLLECTION

The Evaluation Improvement Program

TRANSPARENCIES

Transparencies	Subjects	Trainer's Guide	Evaluator's Guide
TR.E-0	The Evaluation Process	E-2	E-2
TR.E-1	Make Arrangements with School/Program Personnel	E-2	E-2

LEARNING EXERCISES

14	Planning for Data Collection	E-12 - E-14

Introduction

The central purpose of this section is to stress the importance of the data-collection function in the program evaluation process and of careful planning and advance preparation to be certain that all the people involved are ready at appropriate times to carry out their tasks. Things the program evaluator needs to do are covered, as are the monitoring procedures needed to give assurance that data are collected as planned.

Things to Stress in This Section

1. Those points shown in TR. E-1
2. Data collection requires careful management. The best design can be totally destroyed if data are improperly collected.
3. Cooperation in data collection depends on perceived benefits to students, staff, and administration. The more apparent it is to all that the program evaluation will meet their needs, the more confidence one may have that the data collected will be viable.
4. Decide who your data collectors will be and plan a careful inservice training program for them. Even veteran test administrators can benefit from a periodic refresher course. If an instrument is being used for the first time in a school, go over the instrument and directions for its administration in detail with the data collectors. Never rely on written instructions alone.
5. Set schedules which will be free of interruptions.
6. Monitor the collection of data. Check to see that everything is going as planned.

TR. E-0

In the last section, which discussed the selection of instruments, the reliability and validity of the instruments was stressed. No matter how reliable and valid an instrument is, its usefulness can be completely destroyed by carelessness in the collection and handling of data. The need for detailed data collection procedures must be emphasized.

The eight stages of evaluation are interdependent: An evaluation is only as strong as its weakest element. Consider a few of the more obvious faulty data collection problems.

Suggestions for Teaching

1. TR.E-0 Tie data collection to other parts of the evaluation process.
2. TR.E-1 Introduce this section and then have participants scan E-1 through E-6.
3. Use the anecdote on E-6 to initiate a group discussion of similar experiences.
4. Distinguish between methods used to collect data from groups of persons and to collect survey data from individuals.
5. Ask participants to turn to page E-8 while you summarize the checklist for evaluators.
6. Review the sequence of data collectors' activities presented on E-9.
7. Collection of survey data (E-10 and 11). Stress the need for achieving high percent returns and go over methods of obtaining them.
8. Introduce Learning Exercise 14, Planning for Data Collection (E-12 through 14). Have participants work in small groups. Summarize after a short period of work at tables.
9. Give careful attention to monitoring program processes and activities (see E-15 and 16). Point out two versions of forms for recording and monitoring information (E-17 and 18).

EVALUATION TRAINER'S GUIDE

Section F

DATA ANALYSIS

The Evaluation Improvement Program

TRANSPARENCIES

Transparencies	Subjects	Trainer's Guide	Evaluator's Guide
TR. F-0	The Evaluation Process	F-2	F-1
TR. F-1	Types of Statistics	F-2	F-1
TR. F-2	Combining Results	F-3	F-2 and 3
TR. F-3	Frequency and Central Tendency	F-4	F-4
TR. F-4	Number of Persons	F-5	F-5
TR. F-5	Bimodal Distribution	F-5	F-6
TR. F-6	Results of Achievement Test (N=25)	F-5	F-7
TR. F-7	Mean, Median, and Mode in Skewed Distributions	F-5	F-9
TR. F-8	Cases I, II, III: Same Mean Varying Range	F-5	F-9
TR. F-9	Case I: Standard Deviation/ Score Spread	F-5	F-10 and 11
TR. F-10	Case II: Standard Deviation/ Score Spread	F-5	F-10 and 11
TR. F-11	Case III: Standard Deviation/ Score Spread	F-5	F-10 and 11
TR. F-12	Normal Distribution Curve	F-5	F-12
TR. F-13	Distributions Not Overlapping	F-6	F-14
TR. F-14	Distribution Slightly Overlapping	F-6	F-14
TR. F-15	Overlapping Distributions	F-6	F-15
TR. F-16	Large and Small Standard Deviations	F-6	F-15
TR. F-17	Three Variables, Eight Groups	F-9	F-20 and 21
TR. F-18	Ranking from Scores (Two Equal Scores	F-10	F-22
TR. F-19	Ranking from Scores (Three Equal Scores)	F-10	F-22
TR. F-20	Tests for Ordered Data	F-11	F-22 - 24
TR. F-21	Observed and Expected Responses	F-12	F-25
TR. F-22	Significance Levels for Chi Square	F-12	F-26
TR. F-23	Responses on Tax Increase Issue	F-12	F-27
TR. F-24	Calculation of Chi Square	F-12	F-25

Introduction

This section is designed to help workshop participants feel more comfortable with some of the statistical concepts commonly used in the analysis of program evaluation data. A secondary purpose is to emphasize the critical importance of including data analysis as one of the resources that must be used for good program evaluation.

Topics Covered in This Section

- Types of Statistics - Descriptive and Inferential
- Measures of Central Tendency
- Measures of Dispersion or Variability
- Variance and Testing of Differences between two Measures (Hypothesis Testing)
- Non-normal Distributions
- Tests for Category Data
- Tests for Ordered Data
- Tests for Score Data

Things to Stress in This Section

1. Descriptive statistics and inferential statistics serve two different purposes. The presentation of program evaluation data to unsophisticated audiences may rely almost entirely on descriptive statistics, but program evaluation cannot be considered complete so far as the decision maker is concerned without inferential tests that tell you whether the change was real.

2. In the coin-flipping demonstration, it is important to convey the idea that any given measure is an <u>estimate</u> of the true state of affairs. Exactness in measurement is not possible in the behavioral realm. The relationship of this demonstration to the concept of randomness is also important (see F-3).

3. Spend sufficient time on the normal distribution (figure at F-12, text at F-10 and 13). Explain something about the various scales.

4. Be sure you draw the distinction between the normal curve (theoretical model) and how real data distribute themselves. Many data do approximate this model, some do not.

5. The section on variance (F-13 through 15) lays the foundation for an understanding of inferential statistics. Take enough time so participants will get an intuitive grasp of why variance is important.

> This section serves as a brief introduction to statistics and data analysis. We are aware that some of you may have received little formal training in this area. We have prepared the exercises, concepts, and examples with this in mind. Those of you with experience and training in statistical analysis please bear with us. We hardly expect that those not familiar with statistics will learn all about that subject in this session. If you leave feeling more comfortable with the statistical notions expressed here, feeling that you have an intuitive grasp of the concepts, that you have a better idea of what data analysis can buy you in the way of useful information for evaluation, and where you can find it, then we will have achieved our purpose.

Suggestions for Teaching

1. TR. F-0 Tie data analysis to other parts of the evaluation process.

2. TR. F-1 Draw the distinction between descriptive and inferential statistics (F-1).

3. Coin-flipping demonstration (F-2 through 5)
   Rationale:
   The coin-flipping demonstration shows how
   - Data from individuals can be aggregated for a group.
   - Data from groups are combined for grand totals.
   - Data are organized into frequency distributions.
   - Frequency distributions are graphically displayed.
   - Various descriptive statistics may be obtained.

The first two activities (flipping coins individually and recording individual and group results) are carried out by the participants (F-2).

TR. F-2 After participants have flipped their coins and have combined their individual tallies with others at their table, summarize the results.

Combining Results of Coin-Flipping Demonstration

Table No.	Heads	Tails	No. Persons at Table
1			
2			
3			
4			
5			
6			
7			
8			
Total			

Record table total (number of heads, tails, and number of persons at each table).

Sum all columns to get group totals.

Make the point summarized at F-3 on how accuracy increases with increased coin-flipping (or testing).

Make the points about relationship to randomness suggested in the third paragraph at F-3.

TR. F-3 Move to organizing the data systematically. Record frequencies and products.

The easiest way to obtain a frequency distribution for the total group is to ask how many persons got all 10 heads, how many 9s, 8s, etc., count hands, and record these on the transparency.

Calculate mean, median, and mode, record range (see detail below and at F-4 of the Evaluator's Guide).

Value (No. Heads)	Count/ Frequency	Value X Frequency
10		
9		
8		
7		
6		
5		
4		
3		
2		
1		
0		
Total		
Mean ($\bar{x}$)		
Median (Md.)		
Mode (Mo.)		
Range		

TR. F-4 Lead participants through an exercise in which they draw a line graph of the group's own distribution of heads flipped (see F-5 of the Evaluator's Guide).

Central Tendency, Variability, Non-normal Distributions

Suggestions for Teaching (F-6)

1. Present mean, median, and mode.
   TR. F-5 Illustration of bimodal distribution (see F-6)
   TR. F-6 The effect of extreme scores on the mean (see F-7)
   TR. F-7 Skewed distributions (see F-8)

2. Present Range and Standard Deviation.
   TR. F-8 Point out how range may vary even though the mean may remain constant (see F-9).
   TR. F-9} Note relationship between size of standard deviation and spread
   TR. F-10} of scores (see F-10 and 11).
   TR. F-11}

3. TR. F-12 Discuss the normal distribution (see F-12). Point out that 68 percent of the scores always lie between $\pm$ 1 standard deviation from the mean; 95 percent of the scores lie between $\pm$ 2 standard deviations from the mean. This is in the theoretical curve. Real data will approximate these percentages in direct relation to how much the total distribution is like the theoretical curve. Review the percentile equivalents and stanines. The unequal intervals on both scales can be readily seen. Point out the various kinds of standard scores.

Score	Mean	Standard Deviation
Z	0	1
T	50	10
CEEB	500	100
AGCT	100	20
Wechsler	10	3

4. Rationale for Using Normal Curve: If someone questions using the normal distribution with actual distributions whose shapes do not look normal, you can explain the theoretical basis which allows wide usage of the curve, as follows:

> Think of a population with mean M and standard deviation σ (no matter what its distribution). Draw many independent, random samples of size N from the population. Compute the means $\bar{x}_1, \bar{x}_2, \bar{x}_3 \ldots \bar{x}_n$. The distribution of these means will be very nearly normal and closely approaches a normal distribution as more samples are drawn. This distribution of means itself will have the same mean and standard deviation as the original population. The coin-flipping demonstration demonstrates this. This is called the Central Limit Theorem and is the major reason the normal distribution is used.

5. Present variance (see F-13 through 15). This forms the basis of intuitive understanding of inferential tests.

   TR. F-13} Point out that as the amount of overlap increases, our uncertainty
   TR. F-14} about "real" change increases (F-14 and 15).
   TR. F-15}

   TR. F-16 Contrast the amount of overlap (a function of spread or dispersion) for the two distributions with a constant difference between means (F-15).

6. Ask participants to turn to page F-16 and look at other types of distributions as you discuss them.

7. Ask participants to turn to page F-17 as you summarize the various descriptive statistics.

Inferential Statistics (F-18)

Things to Stress in This Section

1. Stress the importance of matching the inferential test to the kind of data at hand. The terms "score," "ordered," and "category" data are substitutes for the terms "interval," "ordinal," and "nominal" data in the mathematics world. They are used here because of the negative reactions received in the initial years of the project by individuals unfamiliar with conventional mathematical terminology.

2. Make it clear that this overview of inferential statistics is not intended to equip participants to use these techniques if they have not had previous training but to offer a speaking acquaintance with them. For those participants who have had an adequate foundation, the treatment is designed to place the various tests of significance in perspective and to serve as a review of their uses.

3. Size up your workshop group. Many persons automatically tune out anything related to quantitative methods. Try to focus on concepts more than calculations. It is usually best not to try to teach computational routines. Occasionally, interested persons with numerical facility will surface. The optional computational sections are designed for individual use by these persons.

4. The tests of significance presented are some of the more frequently used ones, parametric methods for tests of score data, nonparametric methods for tests of ordered data and category data.

5. The same thing can be said about selection of statistical tests as about selection of evaluation designs: Both require attention at a high level of technical knowledge. The exposure the participant receives on inferential statistics here should make him or her better qualified to recognize what skills are needed by data analysis members on a program evaluation team.

Inferential statistics provide a way to test the significance of results obtained when data are collected. As noted in the discussion on descriptive statistics (F-3), all measurement is subject to error (due to inherent difficulties in measuring behavior and to specific testing conditions) and to random fluctuation (due to the particular persons included in the sample being tested). Inferential statistics (F-18) provide a way to separate chance errors and random fluctuation from "real" changes. What is left after inferential statistics have been applied are some answers that can be given with confidence on status or change.

Statistical Tests for Score Data

Things to Stress in This Section

1. Statistical tests for score data are presented only to give exposure to the participant. Convey the concepts, results, and applications.
2. Encourage participants to concentrate on the uses of these tests of statistical significance to help them better select program evaluation team members who have the required skills.

If the data you have can meet the assumptions underlying tests for score data (F-18), there are many different and potentially powerful tests that can be used. Most inferential tests for score data require special training for their proper selection and use. Unless the program evaluator has had this training, he or she is advised to seek the help of someone who has.

Several commonly used tests will be mentioned, but no effort will be made to teach computational routines. Program evaluators who have access to a computer center may wish to seek assistance from that source once the decisions have been made as to what kinds of analyses are appropriate. Do not expect computer people to help you decide what analysis is most appropriate. They may be statisticians as well as data processors, but most are not.

Suggestions for Teaching

1. Discuss t-tests (F-20), stressing the need to know which computational routine to use. There are variations depending on the specific data at hand.

2. Discuss analysis of variance (F-20 and 21). The factorial designs discussed at C-9 through 11 represent one use of analysis of variance. Two-way analysis of variance is like a t-test. But analysis of variance can be used also when the number of groups being compared is three or more.

   TR.F-17 Briefly review expanded designs and factorial possiblities (C-9 through 11). Then talk through F-20 and 21 using TR F-17 as the basis for discussion.

3. Discuss multiple regression (F-21).

Statistical Tests for Ordered Data

Things to Stress in This Section

The four techniques presented here provide a glimpse into the world of non-parametric statistics.* Non-parametric statistics do not make as stringent assumptions about the data as do tests for score data (parametric). Solutions are not given except for the Kruskal-Wallis Test. Other commonly used tests not mentioned are the Mann-Whitney U Test (a method to compare two independent samples) and the Wilcoxon Matched-Pairs Signed-Rank Test (a method to compare two correlated groups). Exposure to a few non-parametric techniques is the objective of this discussion.

The major distinctions to be made when choosing among non-parametric tests of statistical significance are:

a. How many groups are being compared? In general, different tests are used to compare two groups and three or more groups.

---

*Chi Square is also a non-parametric technique (see F-25 through 28).

b. Are the groups independent or are they correlated? When students are randomly assigned to different groups, the groups are considered independent. If more than one set of ordered data is collected on the same students, use statistics for correlated groups.

Ordered data may be obtained in two ways (F-22). First: No numerical scores are obtained, but you are able to place persons or objects along some dimension of interest (as when a committee reviews five textbooks and can make decisions in ordering them on a scale from most to least preferred). Second (and most common): You have obtained test scores but feel they are not precise enough to meet the assumptions for using statistical tests for score data. So you order scores from high to low and use that ranking as ordered data.

Suggestions for Teaching

1. TR. F-18 Demonstrate how to convert score data to ranked data (F-22). TR. F-19
2. Discuss sign test (F-22 and 23). Have participants look at the table on page F-23 as you discuss.
3. Discuss Kruskal-Wallis Test (F-23 and 24). For those who wish it, the solution to this test is given below:

Solution to Kruskal-Wallis Test

$N = 7 + 8 + 9 = 24$

$T = 92.0 + 96.5 + 111.5 = 300$

$$H = \frac{12}{N(N+1)} \Sigma \frac{T_a^2}{N_a} - 3(N+1)$$

Solve for $\Sigma \frac{T_a^2}{N_a}$:

$$\frac{92^2}{7} + \frac{96.5^2}{8} + \frac{111.5^2}{9} = 3754.53$$

Solve for $\frac{12}{N(N+1)}$:

$$\frac{12}{24(25)} = .02$$

Solve for $3(N + 1)$:

$$3(25) = 75$$

$H = .02(3754.53) - 75$

$H = .09$

Refer to χ^2 table of significance levels (F-26).

df = (3 - 1) = 2

This result is not significant.

4. Briefly mention Rank Sums Test and the Friedman Test (F-24).

5. TR. F-20 Tests for Ordered Data (F-24).

Statistical Tests for Category Data

Suggestions for Teaching

1. Introduce Chi-Square Test (F-25): a test of deviation from expected frequencies and a test of association

2. Explain Chi-Square application for testing deviation from expected frequencies and talk through the example (F-25 through 27).

TR. F-21 This first example is based on the following formula

$$\chi^2 = \Sigma \frac{(|f_o - f_e| - .5)^2}{f_e}$$

where f_o = observed frequency
F_e = expected frequency

.5 is a correction for continuity which is recommended by many current statistical treatments of the subject. The correction for continuity is needed because a computed χ^2, being based on frequencies (which are whole numbers), varies in discrete jumps, whereas the χ^2 values given in tables are based on a continous scale.

TR, F-22 Explain again degrees of freedom and levels of significance. Show how to read table of significance levels (F-26).

3. Explain Chi-Square application as a test of association (F-27 and 28).
   TR. F-23 Show how data are displayed for test of association.
4. Present rules which must be followed when using χ^2 (F-28).

Summary on Data Analysis

Suggestions for Teaching

1. Talk through Summary on Inferential Statistics (F-29).
2. Present Data Interpretation Guidelines (F-30 and 31).

An Option

TR, F-24 Computation of χ^2 exercises (F-32 through 35). Be selective. Use only if participants show sufficient interest. If the group is divided, arrange for noninterested participants to take a break or hold a computation session later for those who are interested.

EVALUATION TRAINER'S GUIDE

Section G

REPORTING

The Evaluation Improvement Program

TRANSPARENCIES

Transparencies	Subjects	Trainer's Guide	Evaluator's Guide
TR. G-1	The Evaluation Process: Plan-Conduct-Use	G-1	
TR. G-2	Evaluation Results Should Be Reported	G-1	G-1
TR. G-3	Reporting Evaluation Results	G-1 and 2	
TR. G-4	Interim Report	G-2	G-2
TR. G-5	Final Report	G-2	G-6
TR. G-6	Reporting Results	G-2 and 3	
TR. G-7	Reporting Results	G-3	
TR. G-8	Preparation of the Final Report	G-3	G-6 - 8

LEARNING EXERCISES

16	Recipients and Uses of Interim Evaluation Data	G-4 and 5
17	Determining Appropriate Data Displays	G-9 - 13
18	Writing Recommendations for the Final Report	G-14 - 17
19	Analyzing Program Evaluation Recommendations	G-18 - 20

Introduction

The purposes of this section are to describe the audiences to whom program evaluation results are to be reported, the critical importance of tailoring reporting, and the several methods of doing so. This workshop limits itself to reporting for local purposes; state and federal programs typically have their own evaluation reporting requirements.

TR. G-1

> Let's review where we are now in the content of the workshop. First, we dealt with planning the evaluation and found that careful consideration of questions to be answered in final reports provides assurance that we are gathering the right data. Next, we discussed conducting the evaluation to collect and analyze data. Now, we shall talk about using the evaluation data and getting answers to questions raised at the planning stage.

Suggestions for Teaching

1. TR. G-2 Project transparency while stating that the reporting of evaluation results is a crucially important aspect of the evaluation process. In this section, we shall be considering a number of audiences to whom results should be reported. (See G-1.)

2. TR. G-3 Cover all subheadings of this transparency other than "To Whom?" and ask the group to recall them. As the responses come from the group, ask a participant to record them on a chalkboard or a large piece of paper in front of the room.

3. TR. G-3 Now uncover the "Why?" section on the transparency. For each kind of report-recipient named, record on the chalkboard the reasons that particular recipient needs the report or the uses to which he or she might put it.

   Note that the reasons different audiences need to receive program evaluation reports are, in general, similar to the purposes of program evaluation discussed in Section A. Emphasize that program evaluation reports need to be written in terms of uses to which recipients wish to put them.

Now uncover the "When?" section of the transparency. State that some audiences need to make use of program evaluation reports even while the program is in progress. The needs of others come into play at the end of the program and still others need program evaluation both during and after program activities.

Uncover the "How?" section of the transparency. State that we now need to consider possible formats for both interim and end-of-the-year reporting.

4. TR. G-4 The evaluation report developed while the program is in progress is called the interim report (see G-2). Continue by paraphrasing the information on the transparency and in the text.

5. TR. G-5 The final report is developed at the end of the program and is, among other things, a summary of the program (see G-6). Continue by paraphrasing the information on the transparency and in the text.

6. TR. G-6 State that before viewing reporting formats, it might be well to review certain considerations applicable to all evaluation reporting. Amplify some of the principles cited on the transparency such as:

   - Interim reports must be well-timed so that there is ample opportunity for decision makers to modify the program or the program evaluation while they are in progress.
   - Reports should be clearly and briefly stated. Busy people prefer information that is synthesized and recorded with conciseness.
   - The formats of reports should be easily understood and tailored to each audience.
   - Adequate communication of program evaluation information to all persons participating in the evaluation process is a key to success. Many persons are fearful of program evaluation. If everyone knows how the results are to be used, however, fear may be diminished. Program evaluation should be seen as a vehicle for assisting implementers--including classroom teachers--

improve programs to benefit children. Direct experience with program evaluations will help dispel apprehensions that the results from such evaluations might be used in staff-evaluation procedures. If all persons are involved in one way or another in the evaluation process and know and agree as to the purposes for which the results will be used, their fears will be set aside, especially if reports are well done and are widely useful.

- The presence of fears as noted above should not dissuade an administration from its plans to conduct and report program evaluations.

7. TR. G-7 Make a summary statement on principles of program evaluation reporting.

8. Discuss the Program Management Review Record on G-3. Point out its versatility as both a program-monitoring tool and as an interim evaluation reporting form.

9. Introduce Learning Exercise 16, Recipients and Uses of Interim Evaluation Data (G-4). The purpose of the exercise is to determine the kinds of individuals to whom an interim evaluation report may be relevant and the uses they might wish to make of the data reported. Discuss after 8-10 minutes.

10. In end-of-the-year reporting, it is important to keep in mind the need to report to a variety of audiences in a variety of styles and to consider appropriate formats for each type of audience.

11. Introduce Learning Exercise 17, Determining Appropriate Data Displays (G-9 through 13). Discuss after approximately 10 minutes.

12. TR. G-8 Uncover topic headings one at a time and discuss the purpose of each as described on G-6 - 8 of the Guide.

13. Reinforce this discussion by asking the participants to read G-6 through 8 on a suggested outline for the final report (three minutes). Follow the reading with a discussion of preparation, review, comment, revision, and approval procedures:

    - Who would likely draft the program evaluation report?

- How might one verify the evaluation findings with the program participants?
- Who might participate in the review process?
- What steps would lead up to issuing the approved report?

14. Introduce Learning Exercise 18, Writing Recommendations for the Final Report (G-14 through 17). As the groups are reading the materials, distribute transparencies or large sheets of paper and felt pens to each group. After the participants have had approximately 15 minutes for discussion, ask each group to record its recommendations to present to the class. Call upon spokespersons from two or three groups to report their recommendations to the total group. After each report, ask the question: "Are those recommendations supported by the data?"

15. State that conclusions and recommendations in the program evaluation report give decision makers bases for action. Then ask the participants to read the instructions for Learning Exercise 19, Analyzing Program Evaluation Recommendations (G-18 through 20). Each participant should rate each of the recommendations and then compare the results of his analysis with those of another participant. Encourage two-person teams to discuss the differences they may have. Conclude after approximately 15 minutes.

16. State that clear, concise reporting is essential to the next step, applying program evaluation findings.

EVALUATION TRAINER'S GUIDE

Section H

APPLYING FINDINGS

The Evaluation Improvement Program

TRANSPARENCIES

Transparencies	Subjects	Trainer's Guide	Evaluator's Guide
TR. H-0	The Evaluation Process	H-1	
TR. H-1	Summary of Steps	H-1	
TR. H-2	There is an Alternative	H-2	H-1

LEARNING EXERCISES

20	Use of Evaluation Information	H-2 - 11
21	Roadblocks to Program Evaluation	H-12

Introduction

The purpose of this section is to reinforce the point that has been made throughout the Guide: that evaluation serves no purpose until it is used. The primary activity is an exercise designed to provide experience in analyzing a program evaluation report from the perspective of different audiences and to try to make explicit some of its practical uses.

Suggestons for Teaching

1. TR. H-0 Integrate the eight elements of program evaluation and focus upon applications and uses to be made of the results.
2. TR. H-1 Summarize the steps to be taken to accommodate each audience. The purposes of these steps is to give the evaluator direction for the evaluation plan, to sensitize each audience to its own needs and responsibilities, and to make all more aware of potential uses of evaluation information even before the program begins.
3. Introduce Learning Exercise 20, Use of Evaluation Information (H-2 through 11). Have participants work in groups, arranging with each one to report the findings of his or her group during the discussion period. You may have each group prepare a transparency or report from notes, whichever seems most appropriate. The most important part of the exercise is the discussion following the several group reports. Here are some questions you may wish to ask:
    - What audiences do you think this report would be appropriate for? Give examples of why you think so.
    - What questions do you think the program evaluation might have addressed?
    - Are there parts of the program evaluation plan that you would quarrel with? What would you have done differently?
4. Discuss the need for strategic planning to anticipate roadblocks to effective use of program evaluation results and neutralize them. Introduce

Learning Exercise 21, Roadblocks to Program Evaluation (H-12). List all roadblocks identified. Discuss ways to overcome, circumvent, and minimize each obstacle identified.

5. TR. H-2 Conclude the workshop with an optimistic statement on the potential of program evaluation in leading to improvements in a program, in the process of educational administration, and in education itself.